HEPData
(hepdata.net)

Love Data Week 2025
Ghent University

Graeme Watt – Project Manager/Lead

Jordan Byers – Research Software Engineer **(Me)**

What is HEPData?

- **HEPdata** (High Energy Physics) is an online open-access repository for tabular high-level data, containing more than 10K HEP publications (130K data tables)
- Funded by the **UK STFC** (Science & Technology Facilities Council, based in the **IPPP** (Institute for Particle Physics Phenomenology) at **Durham University**
- We apply the **FAIR** principles (Findable, Accessible, Interoperable, Reusable)
- Infrastructure provided by **CERN IT**, with support from **CERN SIS**

Website: https://hepdata.net

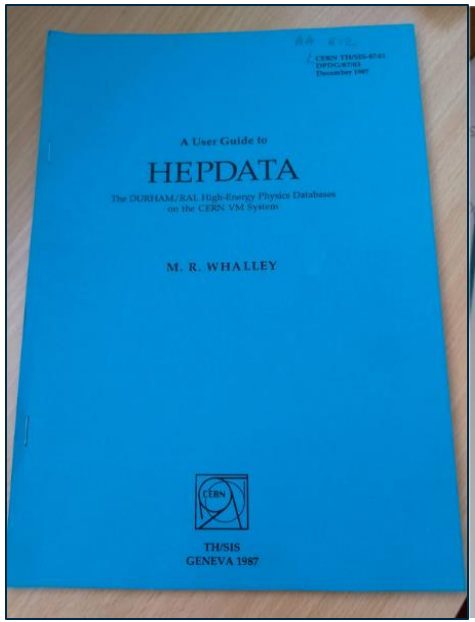
Docs: https://hepdata.readthedocs.io/en/latest/

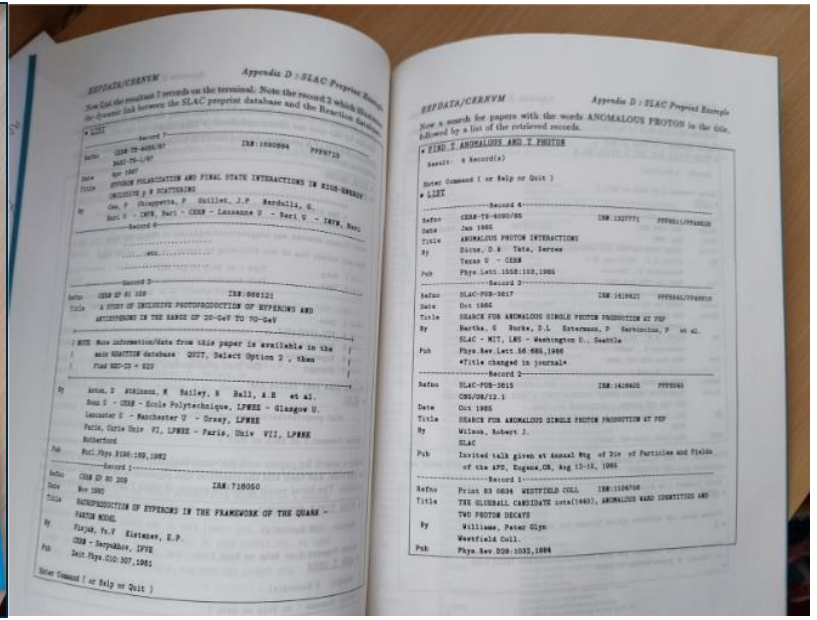
GitHub: https://github.com/HEPData/

HEPData - A Brief History

- Started life in the 70s
- Web interface introduced in the early 90s
- Hosted at hepdata.cedar.ac.uk, until a 2017 redevelopment
- Hepdata.net launching in 2017 in its current iteration

A copy of the user guide from 1987!



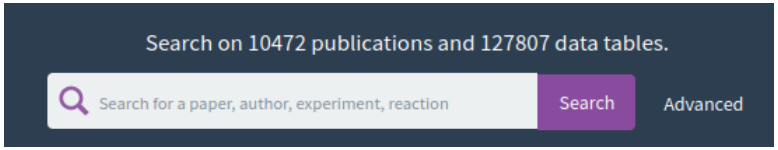


https://cds.cern.ch/record/184048/files/CM-P00063894.pdf

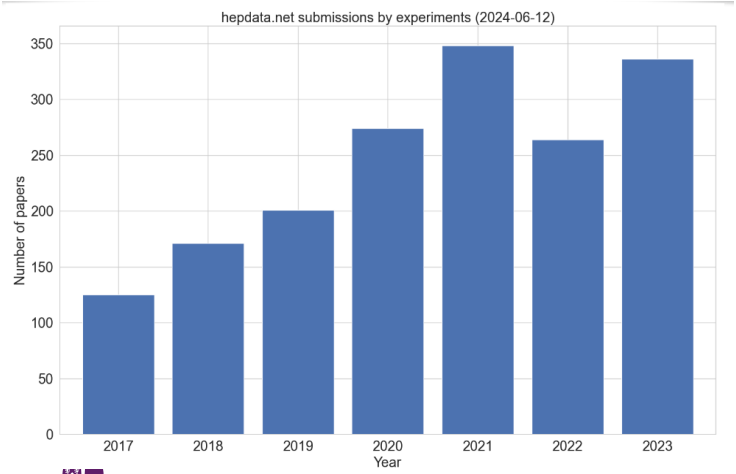
Durham University

HEPData - Homepage

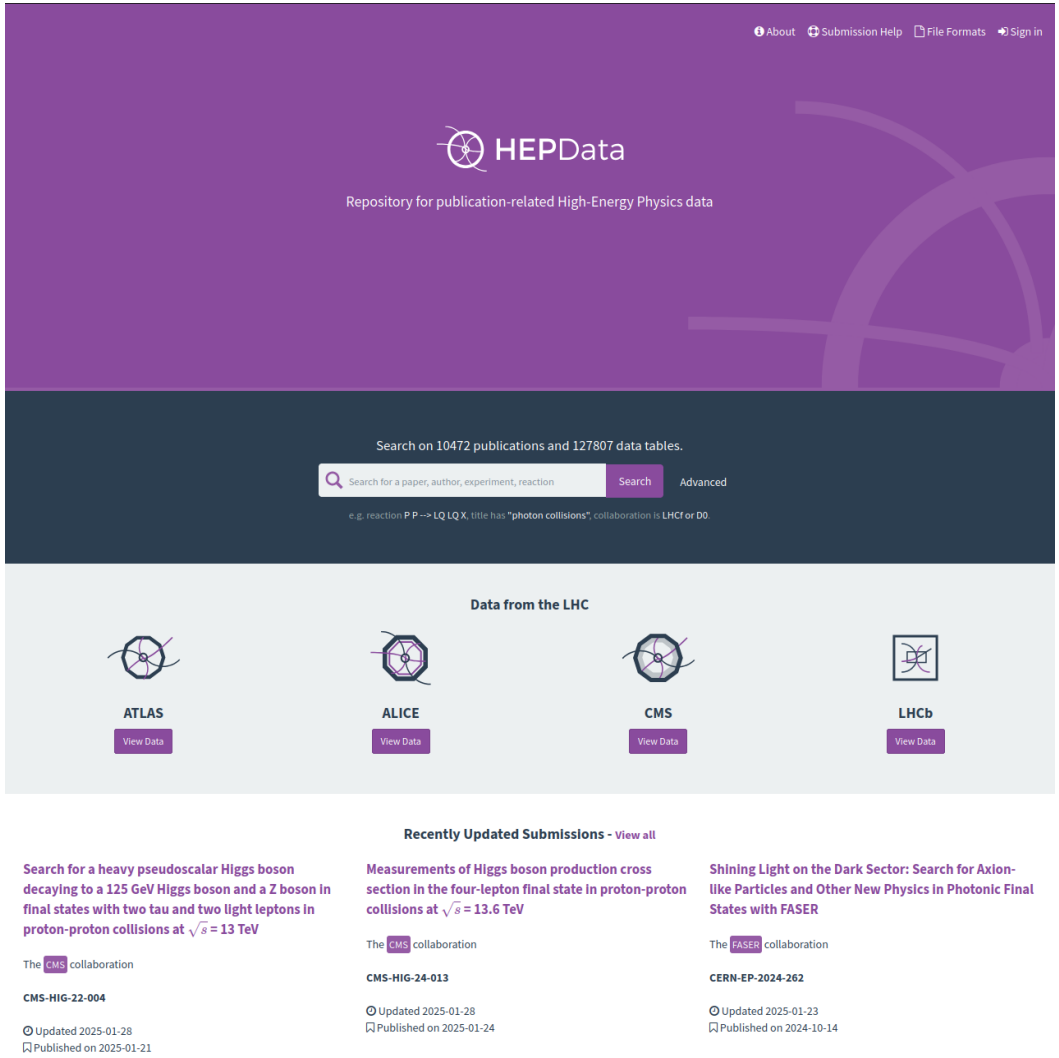
Currently at nearly 10.5k publications



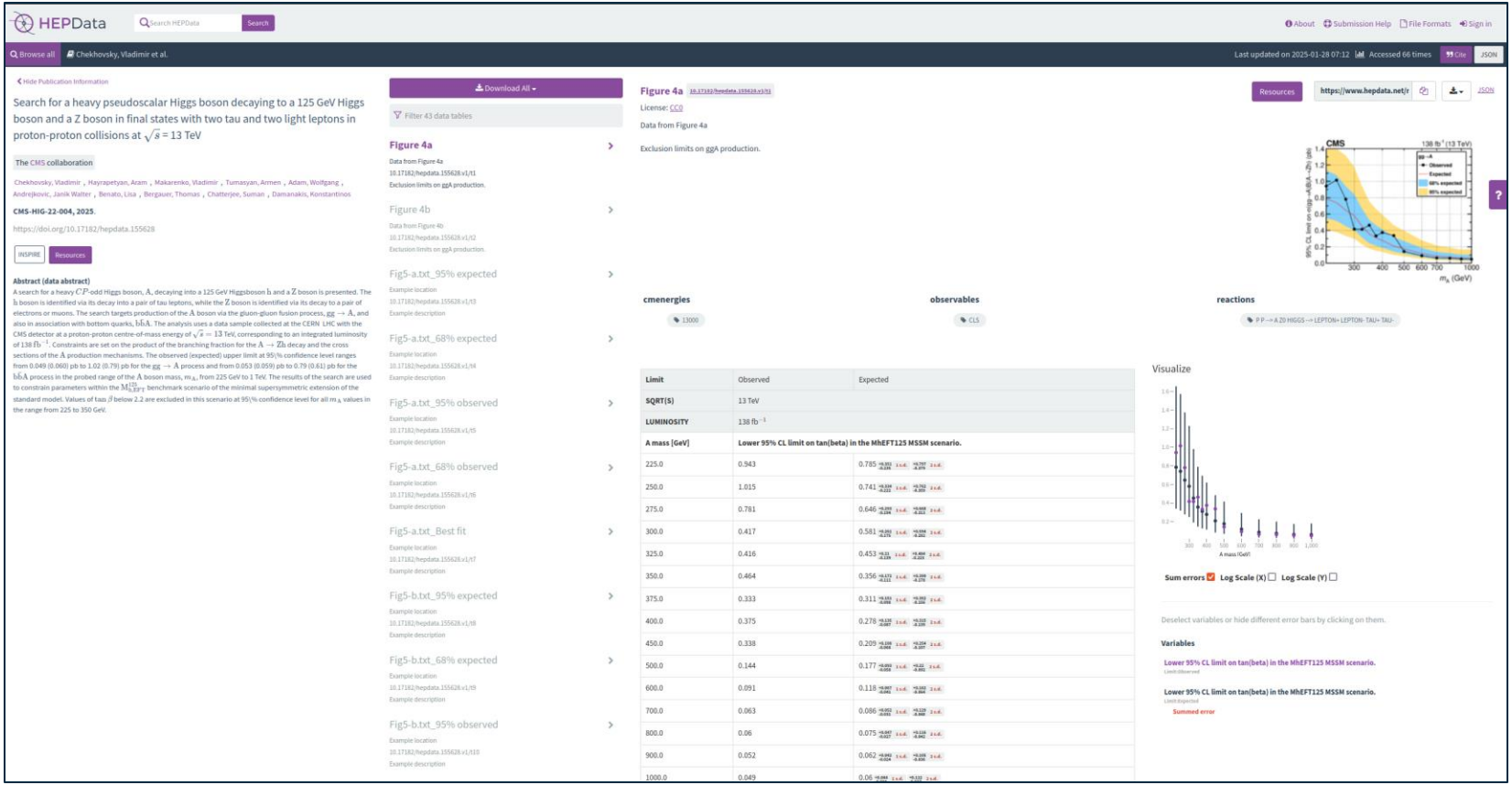
Submissions pear year 2017-2023



Durham University



HEPData - Data

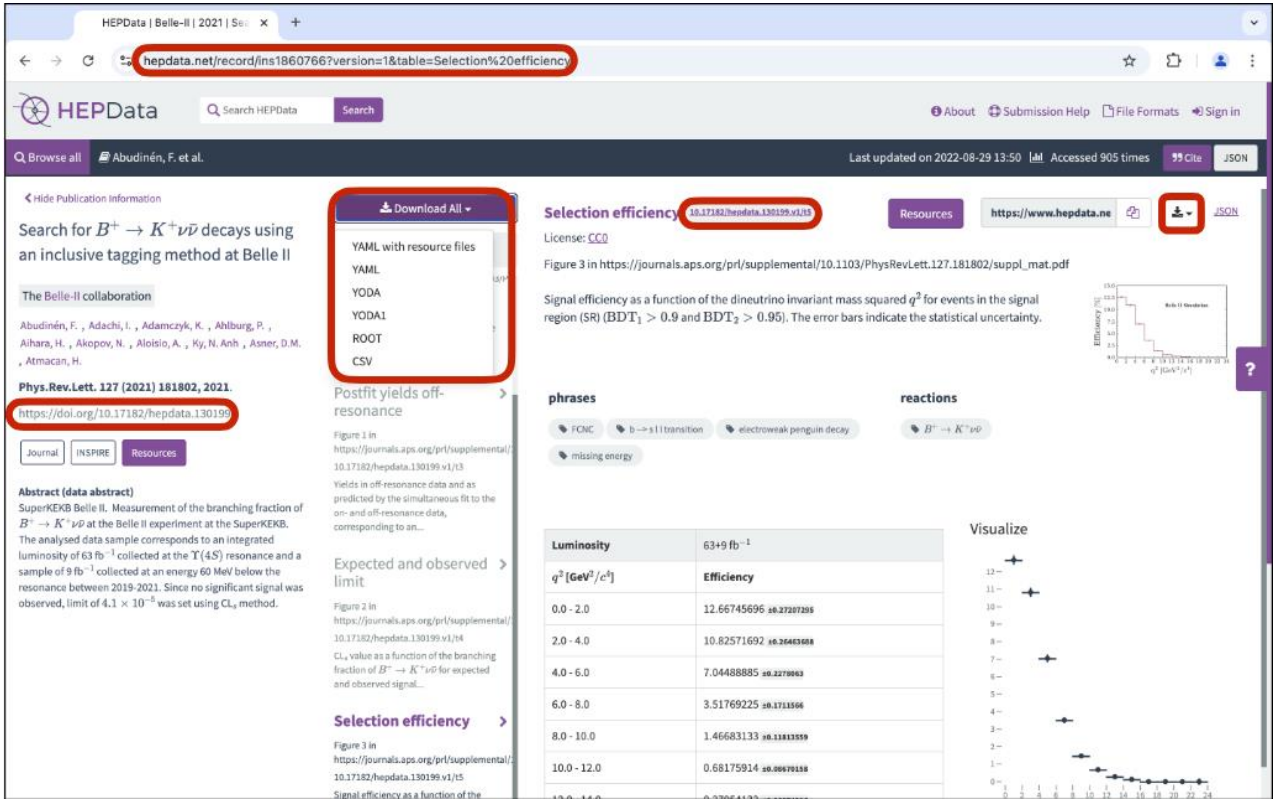


Durham University

https://www.hepdata.net/record/ins2872775

HEPData - Data

- Citable (DOIs)
  - Submissions
  - Data tables
- Downloadable
  - Different formats
- Versioned



Durham University

HEPData – Data (Submission)

Submission information:

- Title
- Experiment
- Authors
- Publication
- DOI
- INSPIREHEP Link
- Resource Files
- Abstract

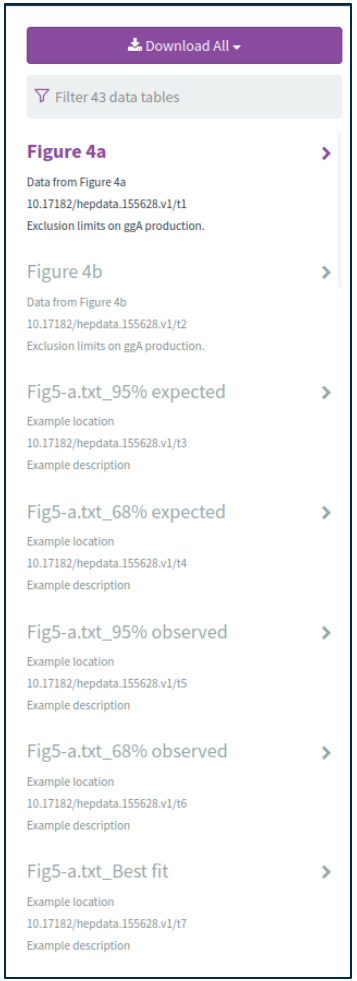
Durham University

https://www.hepdata.net/record/ins2872775

HEPData – Data (Data Tables)

A searchable list of the submission figures

- Download link
- Title
- Location
- DOI
- Description



https://www.hepdata.net/record/ins2872775

Durham University

HEPData – Data (Data Table)

An individual data table:

- Data table information
  - Name, description, license etc.
- DOI
- Resource Links
- Alternate format links
- Visualisations
- Graph Images
- Data Table

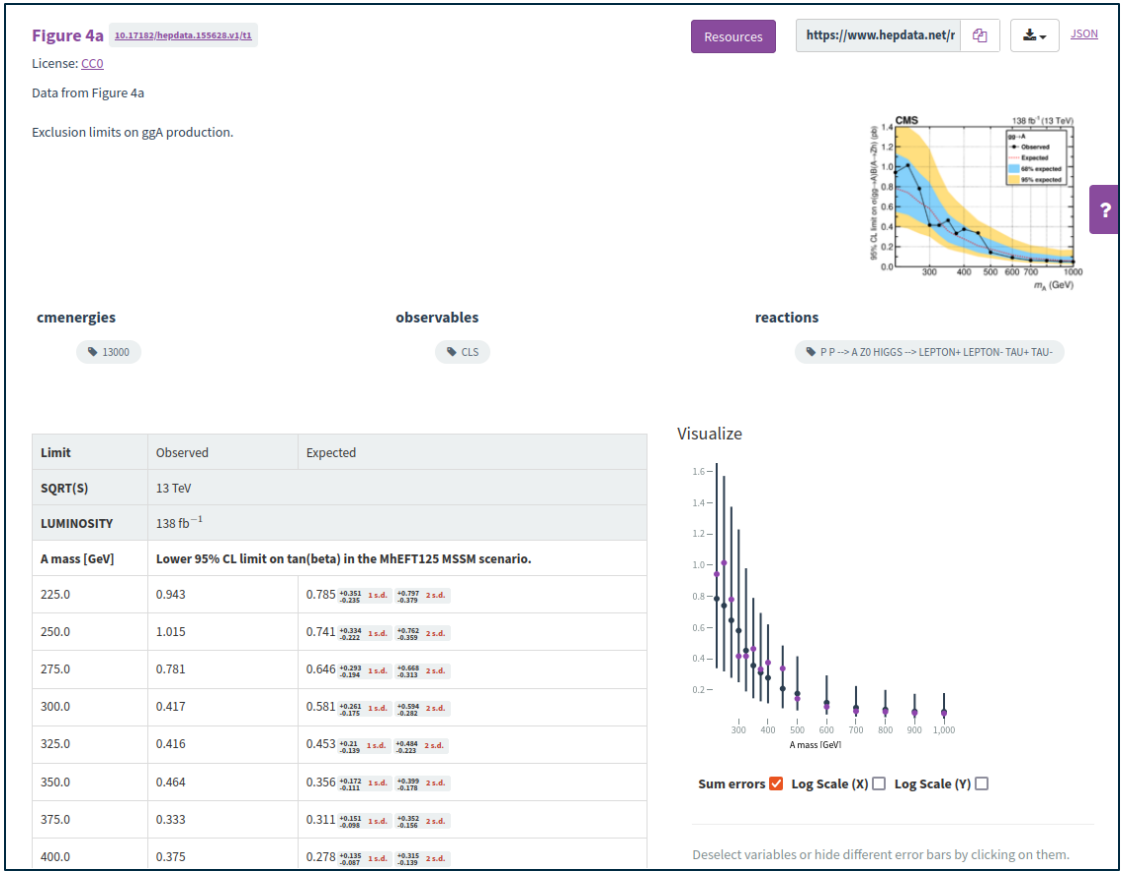


Figure 4a 10.17182/hepdata.155628.v1/t1

Resources | https://www.hepdata.net/r | JSON

License: CC0

Data from Figure 4a

Exclusion limits on ggA production.

cmenergies: 13000

observables: CLS

reactions: P P --> A Z0 HIGGS --> LEPTON+ LEPTON- TAU+ TAU-

Limit	Observed	Expected
SQRT(S)	13 TeV	
LUMINOSITY	138 fb^{-1}	
A mass [GeV]	Lower 95% CL limit on tan(beta) in the MhEFT125 MSSM scenario.	
225.0	0.943	0.785 $^{+0.351}_{-0.235}$ 1 s.d. $^{+0.797}_{-0.379}$ 2 s.d.
250.0	1.015	0.741 $^{+0.334}_{-0.222}$ 1 s.d. $^{+0.762}_{-0.359}$ 2 s.d.
275.0	0.781	0.646 $^{+0.293}_{-0.194}$ 1 s.d. $^{+0.668}_{-0.313}$ 2 s.d.
300.0	0.417	0.581 $^{+0.261}_{-0.175}$ 1 s.d. $^{+0.594}_{-0.282}$ 2 s.d.
325.0	0.416	0.453 $^{+0.21}_{-0.139}$ 1 s.d. $^{+0.484}_{-0.223}$ 2 s.d.
350.0	0.464	0.356 $^{+0.172}_{-0.111}$ 1 s.d. $^{+0.399}_{-0.178}$ 2 s.d.
375.0	0.333	0.311 $^{+0.151}_{-0.098}$ 1 s.d. $^{+0.352}_{-0.156}$ 2 s.d.
400.0	0.375	0.278 $^{+0.135}_{-0.087}$ 1 s.d. $^{+0.315}_{-0.139}$ 2 s.d.

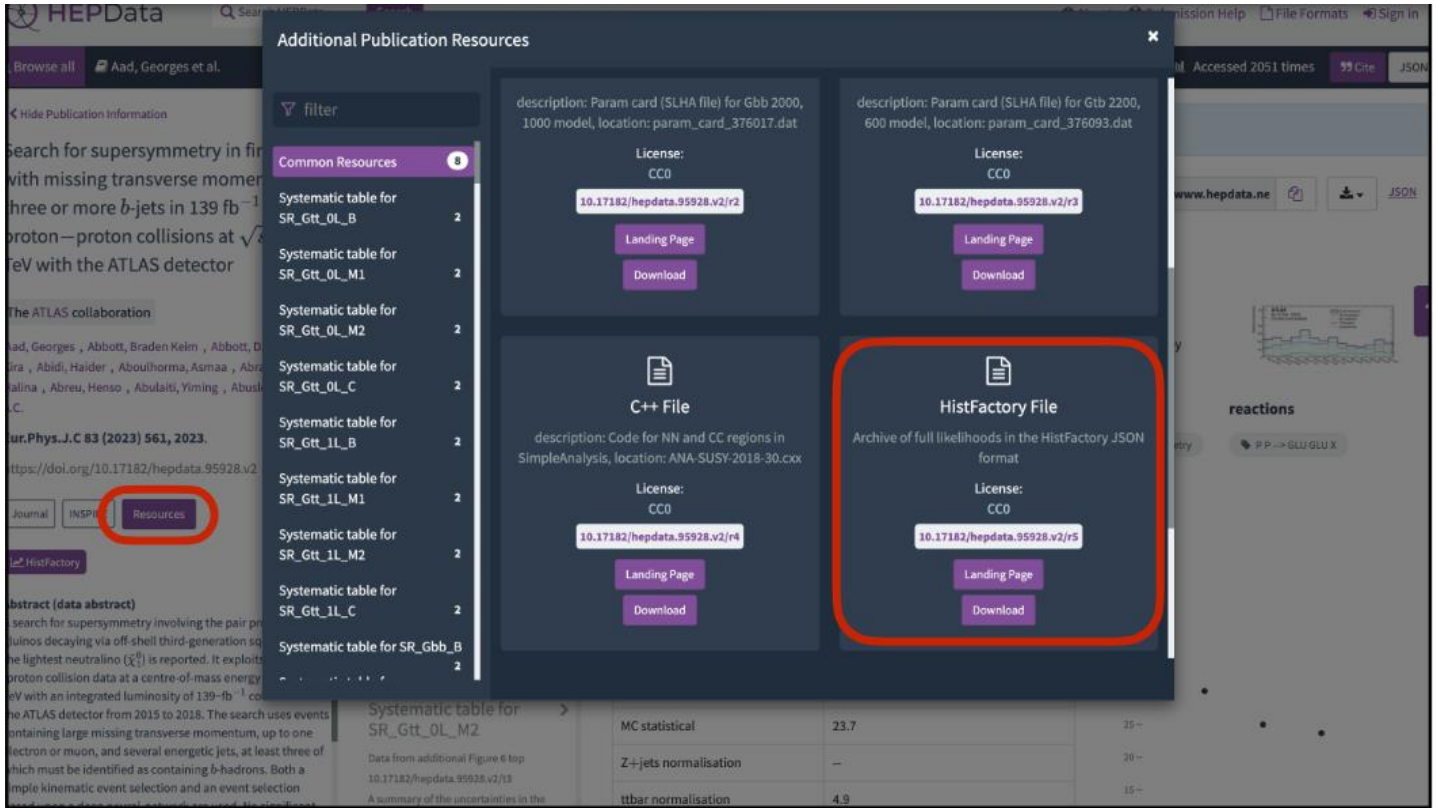
Visualize

Sum errors ☑ Log Scale (X) ☐ Log Scale (Y) ☐

Deselect variables or hide different error bars by clicking on them.

https://www.hepdata.net/record/ins2872775

HEPData – Additional Resources

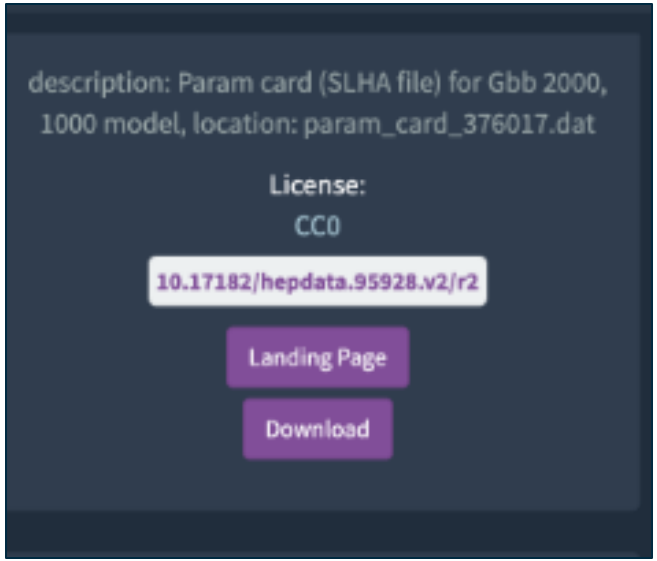


Durham University

https://hepdata-submission.readthedocs.io/en/latest/

HEPData – Additional Resources

These can be accessed by clicking the associated *Landing Page* or *Download* buttons.



Durham University

https://hepdata-submission.readthedocs.io/en/latest/analyses.html

HEPData – Additional Resources (Landing Page)

HEPData | Search HEPData | Search | About | Submission Help | File Formats | Sign in

Browse all | View all Publication Data | png File from Version 1 of Chekhovsky, Vladimir et al. | Last updated on 2025-01-28 07:09 | Cite | JSON

Hide Publication Information

Measurements of Higgs boson production cross section in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13.6$ TeV

The CMS collaboration

Chekhovsky, Vladimir , Hayrapetyan, Aram , Makarenko, Vladimir , Tumasyan, Armen , Adam, Wolfgang , Andrejkovic, Janik Walter , Benato, Lisa , Bergauer, Thomas , Chatterjee, Suman , Damanakis, Konstantinos

CMS-HIG-24-013, 2025.

INSPIRE | Resources

Abstract (data abstract)

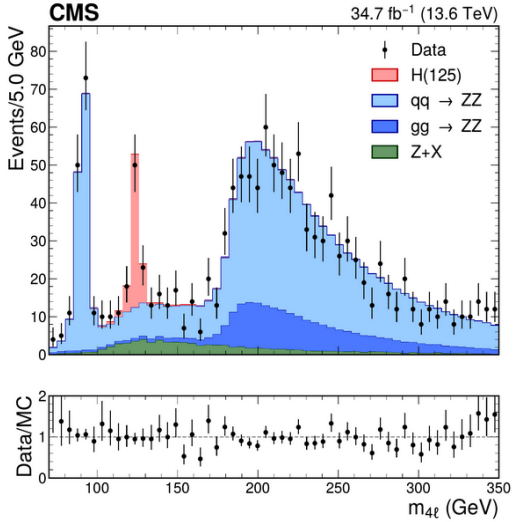
CERN-LHC. Measurements of Higgs boson production cross sections in the four-lepton final state at 13.6 TeV.

CMS-HIG-24-013_Figure_001-a.png

https://www.hepdata.net/r

License: CC0

Image file



Download

https://hepdata-submission.readthedocs.io/en/latest/analyses.html

Durham University

HEPData – Data Formats

Input Format

YAML – Choice for native HEPData format.

```yaml
name: "Table 1"
location: Data from Page 17 of preprint
description: The measured fiducial cross sections.  The first
keywords: # used for searching, possibly multiple values for
  - {name: reactions, values: [P P --> Z0 Z0 X]}
  - {name: observables, values: [SIG]}
  - {name: cmenergies, values: [7000.0]} # centre-of-mass ene
  - {name: phrases, values: [Inclusive, Integrated Cross Sect
data_file: data1.yaml
data_license: # (optional) you can specify a license for the
  name: "GPL 2"
  url: "url for license"
  description: "Tell me about it. This can appear in the main
additional_resources: # (optional)
  - location: "analysis_script.py"
    description: "Analysis script"
  - location: "http://github.com/HEPData/hepdata"
    description: "Full source code for creating this data"
  - location: "root_file.root"
    description: "Some file"
    license: # (optional)
      name: 'GPL 2'
      url: "url for license"
      description: "Tell me about it. This can appear in the
```

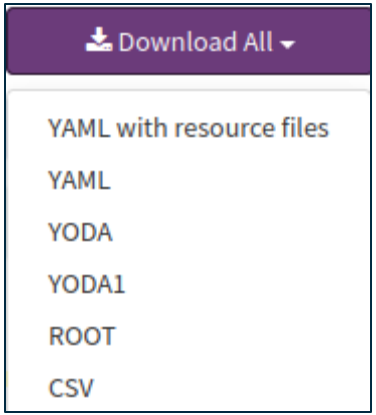
Output Format

JSON - JavaScript Object Notation

CSV - Comma-Separated Values

ROOT - Binary .root File

YODA - For inclusion in a Rivet analysis. (YODA2 and Legacy YODA)



https://www.hepdata.net/formats

Durham University

HEPData - Licences

- HEPData Terms of Use:
  Unless specified otherwise for selected datasets, all metadata and datasets in the HEPData service are made available under the terms of CC0.

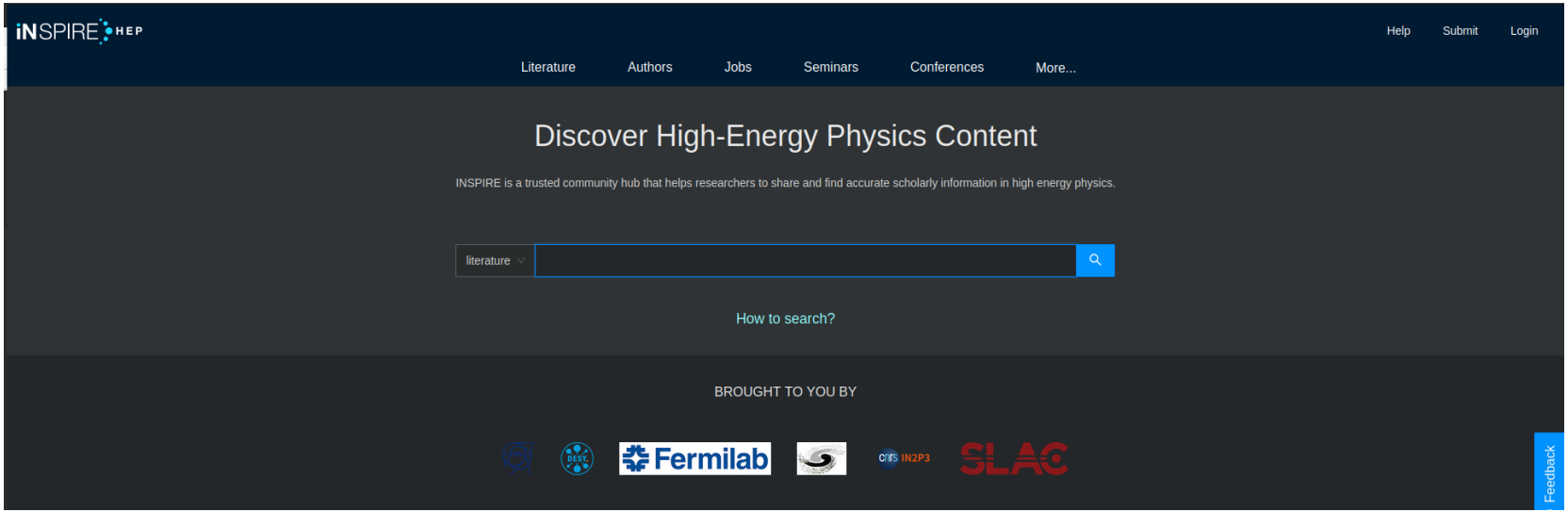
CC0 (aka CC Zero) is a public dedication tool, which enables creators to give up their copyright and put their works into the worldwide public domain. CC0 enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, with no conditions.

NOTE: The user is free to specify a different license if desired.

https://www.hepdata.net/terms

HEPData – InspireHEP

- Main HEP literature platform - https://inspirehep.net/
- You must have an Inspire ID to finalise your HEPData submission
- HEPData submission is linked to an InspireHEP submission



Durham University

https://inspirehep.net

HEPData – Search

Search by text, phrase, keyword, abstract content etc. and using query strings

Search on title, abstract, or record abstract

Find all data with *collisions* in the **title**
title:collisions

Find all data with *"baryon production"* in the **abstract**
abstract:"baryon production"
Quotes force a full match.

Find all data with *"CERN-LHC"* in the **data abstract**
(i.e. the *"comment"* in the *submission.yaml* file)
data_abstract:"CERN-LHC"
Quotes force a full match.

Search by keywords

Find all data with the **phrase** *diffractive* or *elastic*
phrases:diffractive OR phrases:elastic

Find all data with the *P P --> LQ LQ X* **reaction**
reactions:"P P --> LQ LQ X"

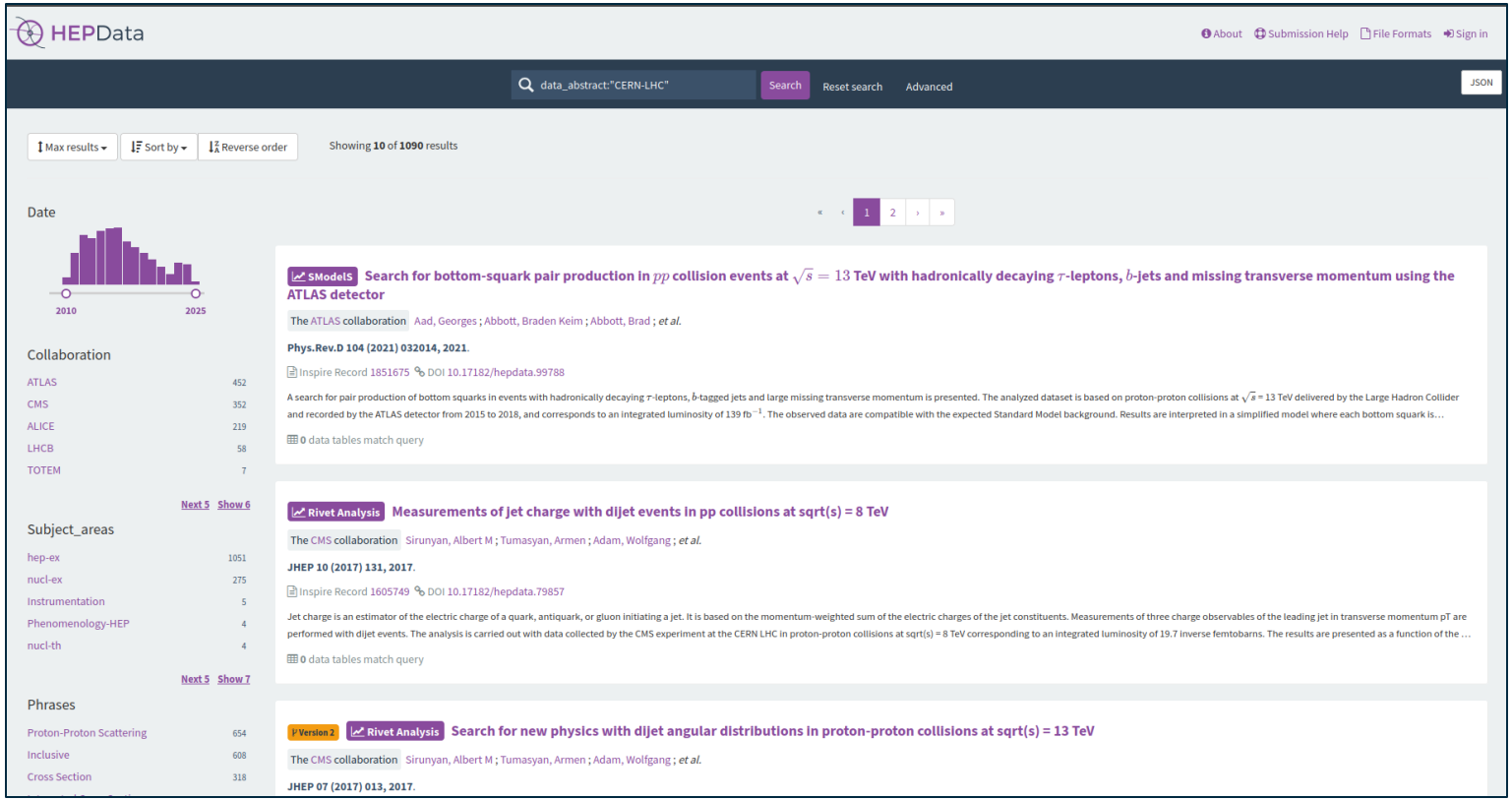
Find data by **CM Energy** (in GeV):
cmenergies:1.34
cmenergies:[1.3 TO 1.4] (inclusive range, i.e. $1.3 \leq \sqrt{s} \leq 1.4$)
cmenergies:{1.3 TO 1.4} (exclusive range, i.e. $1.3 < \sqrt{s} < 1.4$)
cmenergies:[1.3 TO 1.4} (half-open range, i.e. $1.3 \leq \sqrt{s} < 1.4$)

Find all data with **observable** *ASYM*
observables:ASYM

Durham University

https://www.hepdata.net/search/

HEPData - Search



Durham University

https://www.hepdata.net/search/?q=data_abstract:%22CERN-LHC%22

HEPData Submission

Durham University

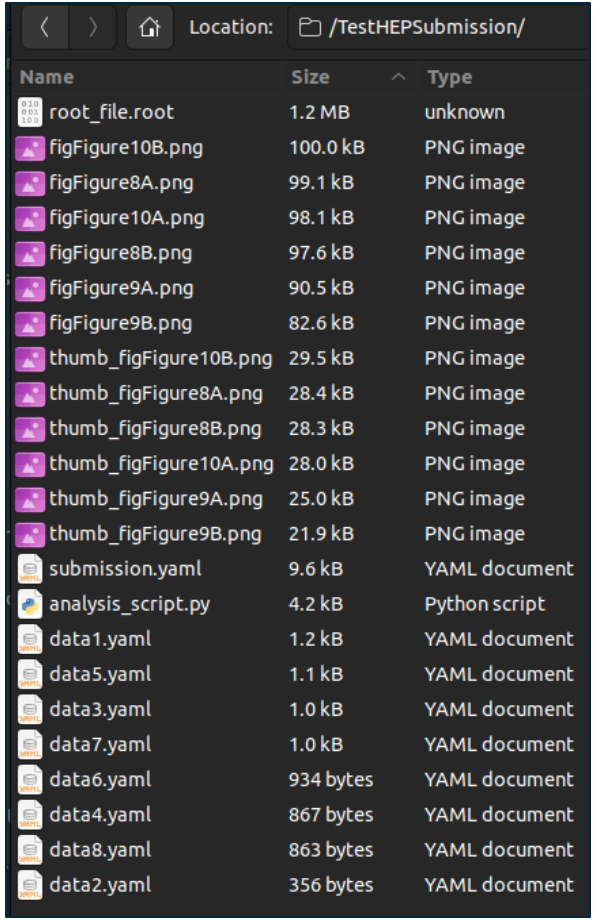
HEPData - Submission

TestHEPSubmission

A HEPData submission is a folder containing the following items:

- **Submission.yaml**
  - This contains core submission data, such as the abstract, additional resource file definitions, other metadata. Also contains a definition for each table, and any associated images/files.
- **Figure images**
  - Any images required for the data
- **Data files (.yaml)**
  - Files containing the data values.
- **Additional resources**
  - Additional files of use for the submission

TestHEPSubmission.zip



Durham University

https://hepdata-submission.readthedocs.io/en/latest/

- https://hepdata-submission.readthedocs.io/en/latest/submission_yaml.html

HEPData - Submission
What is YAML?

Our choice as native HEPData format, is a human-readable data format for data serialisation. Here we map it to our data structure. Uses tab intentation as a part of its formatting.

https://yaml.org/

```yaml
name: "Table 1"
description: Describe the data.  The more you say, the easi
keywords: # used for searching, possibly multiple values fo
- {name: reactions, values: [P P --> Z0 Z0 X]}
- {name: observables, values: [SIG]}
- {name: cmenergies, values: [7000.0]}
- {name: phrases, values: [Inclusive, Integrated Cross Sect
data_file: data1.yaml
```


https://yaml.org/

HEPData – Submission submission.yml

Contains core submission information, including the abstract and any associated additional resources.

Also contains information about submission data files. Most importantly, their location.

Example of the submission-specific data within submission.yaml (Above table information)

```yaml
# Start a new submission. This section is optional for the provision
---
additional_resources: # additional references (e.g. experiment TWiki
- {location: "http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/S
- location: "Likelihoods.tar.gz"
  description: "Archive of full likelihoods in the HistFactory JSON
  type: "HistFactory" # (optional) currently supports 'HistFactory'

comment: | # Information that applies to all data tables.
  CERN-LHC.  Measurements of the cross section for ZZ production usi

---
```

Durham University

https://hepdata-submission.readthedocs.io/en/latest/

HEPData Submission - submission.yml

There is a separate YAML "document" for each data table.

Note: A ***document*** is separated by the "---" indicator within the document.

Schema:
https://github.com/HEPData/hepdata-validator/blob/main/hepdata_validator/schemas/1.1.1/submission_schema.json

An example definition of a table within the submission.yaml file:

```yaml
---
# This is Table 3.
name: "Table 3"
location: Data from Figure 8A
description: Normalized ZZ fiducial cross section (multiplied by 10^6 for
keywords: # used for searching, possibly multiple values for each keyword
- {name: reactions, values: [P P --> Z0 Z0 X]}
- {name: observables, values: [DSIG/DPT]}
- {name: cmenergies, values: [7000.0]}
- {name: phrases, values: [Inclusive, Single Differential Cross Section,
data_file: data3.yaml
additional_resources:
- {description: Image file, location: figFigure8A.png}
- {description: Thumbnail image file, location: thumb_figFigure8A.png}
```

https://hepdata-submission.readthedocs.io/en/latest/

A HEPData Submission – Data Files

Data files, usually called *dataX.yml* contain the data values for an individual data table.

Can be encoded as either YAML or JSON

Defined in two parts:

1. Independent Variables *(X Axis)*
2. Dependent Variables *(Y Axis)*

> a. the *independent variables* (e.g. the x-axis of a plot);
> b. the *dependent variables* (the thing you're measuring, e.g. the y-axis of a plot).

data1.yaml

```yaml
independent_variables:
- header: {name: Leading dilepton PT, units: GEV}
  values:
  - {low: 0, high: 60}
  - {low: 60, high: 100}
  - {low: 100, high: 200}
  - {low: 200, high: 600}
dependent_variables:
- header: {name: 10**6 * 1/SIG(fiducial) * D(SIG(fiducial))/DPT, uni
  qualifiers:
  - {name: RE, value: P P --> Z0 < LEPTON+ LEPTON- > Z0 < LEPTON+ LE
  - {name: SQRT(S), units: GEV, value: 7000}
  values:
  - value: 7000
    errors:
    - {symerror: 1100, label: stat}
    - {symerror: 79, label: 'sys,detector'}
    - {symerror: 15, label: 'sys,background'}
  - value: 9800
    errors:
    - {symerror: 1600, label: stat}
    - {symerror: 75, label: 'sys,detector'}
    - {symerror: 15, label: 'sys,background'}
  - value: 1600
    errors:
    - {symerror: 490, label: stat}
    - {symerror: 41, label: 'sys,detector'}
    - {symerror: 2, label: 'sys,background'}
  - value: 80
    errors:
    - {symerror: 60, label: stat}
    - {symerror: 2, label: 'sys,detector'}
    - {symerror: 0, label: 'sys,background'}
```

https://hepdata-submission.readthedocs.io/en/latest/data_yaml.html

A HEPData Submission – Additional Resources

Other files:

- **Code:** .py/.cpp etc. Files
- **Analyses:** Files containing analysis results
- **Data:** .dat and other formats
- **URLs:** External resource links

Durham University

https://hepdata-submission.readthedocs.io/en/latest/analyses.html

HEPData Submission – Keywords

There are also some OPTIONAL keywords available for your data:

Current keywords are:

- cmenergies (GeV)
- observables
- phrases
- reactions

Example keyword definition within submission.yml

```
keywords: # used for searching, possibly multiple values for each keyw
- {name: reactions, values: [P P --> Z0 Z0 X]}
- {name: observables, values: [SIG]}
- {name: cmenergies, values: [7000.0]}
- {name: phrases, values: [Inclusive, Integrated Cross Section, Cross
```

Durham University

https://hepdata-submission.readthedocs.io/en/latest/keywords.html

HEPData Submission – Extra

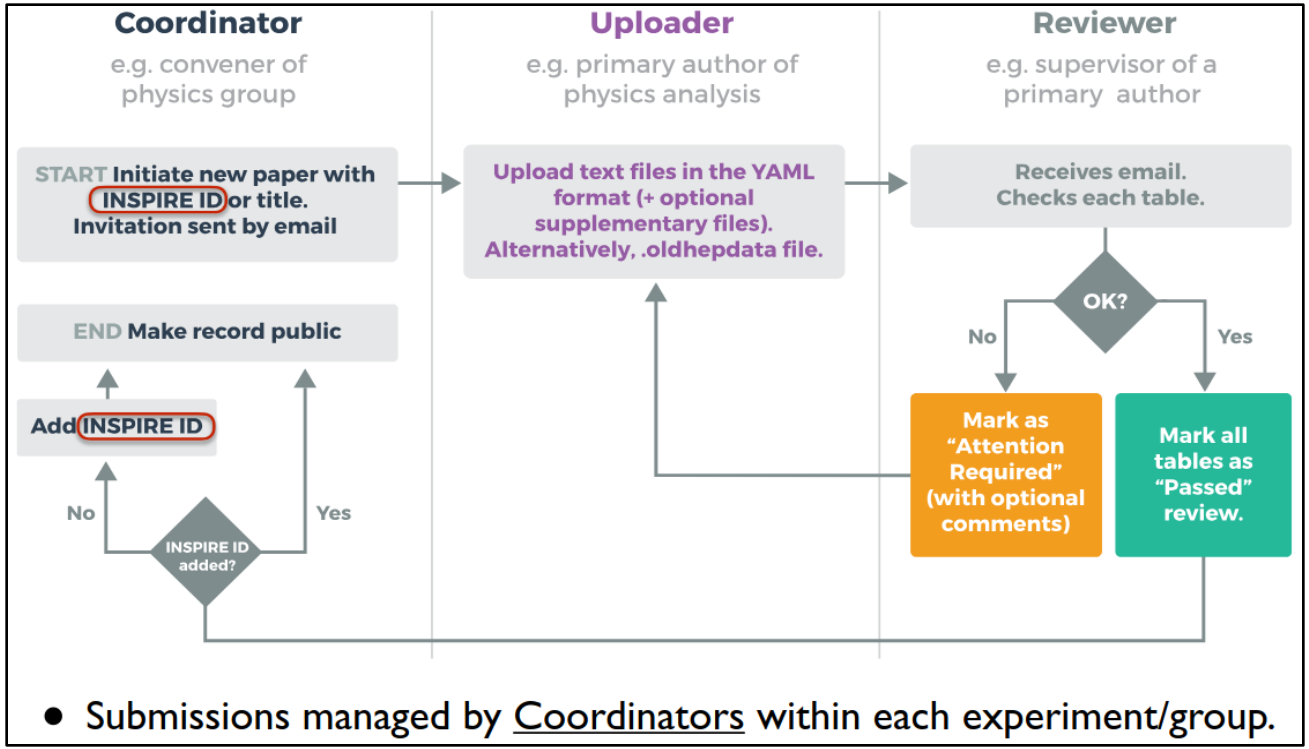
- You must already have an INSPIRE ID to finalise your submission. You can do the review process with only a provisional title.
- Your experiment group must have a coordinator, see https://www.hepdata.net/permissions/coordinators to see if your experiment group already has one, or you can apply to be one through your HEPData Dashboard.
- All HEPData YAML files are checked against the JSON schema contained in the hepdata/hepdata-validator repository.

https://hepdata-submission.readthedocs.io/en/latest/

HEPData
Creating a Submission

Durham University

HEPData - Submission Process

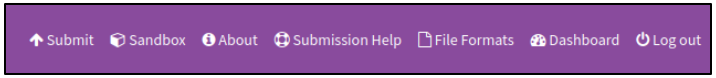


- Submissions managed by Coordinators within each experiment/group.

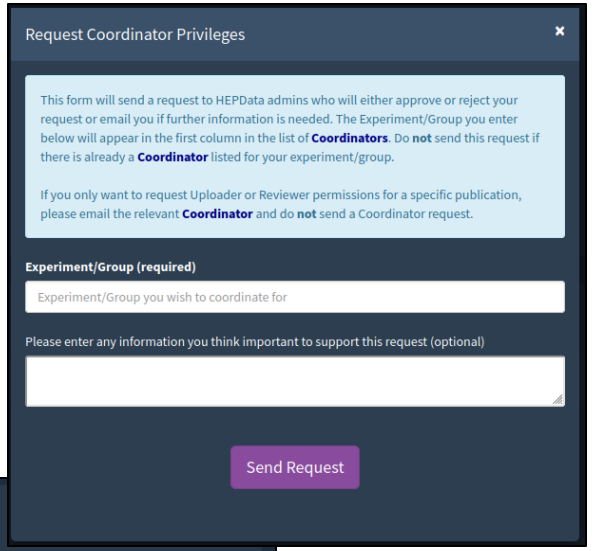
Durham University

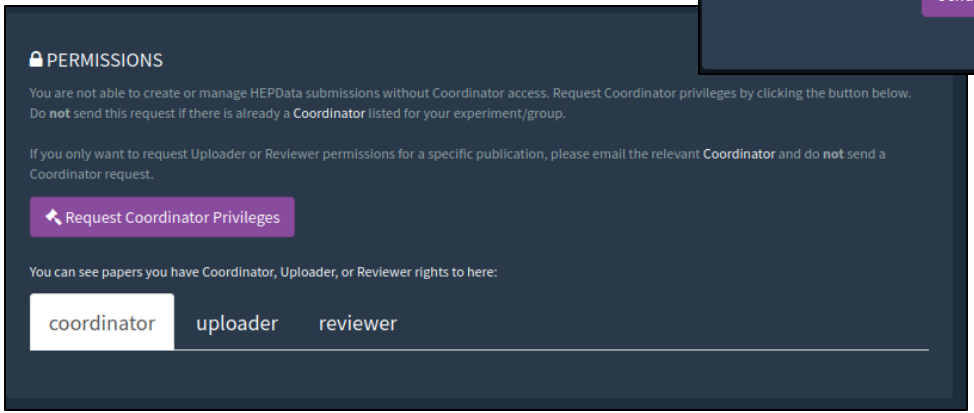
HEPData – Creating a Submission (As coordinator)

To create a new submission, use the "Submit" button on the top right (if you are a coordinator), of the HEPData.net webpage, which is located on the top toolbar.



If you do not see the *Submit* button, then you need to become a coordinator. You can do this through the *Dashboard*.



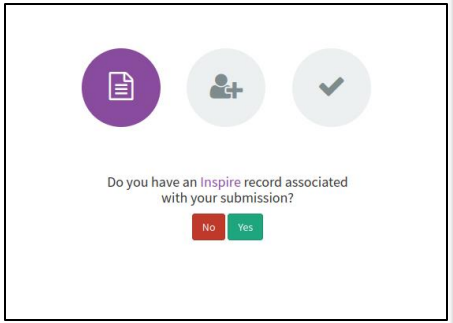


Durham University

https://ww.hepdata.net/permissions/coordinators

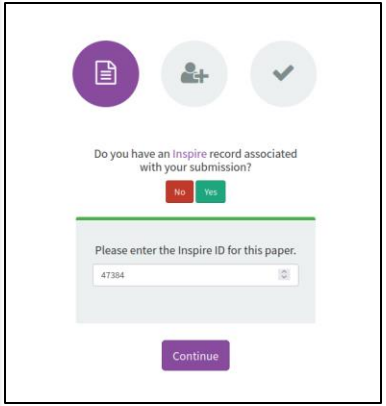
HEPData – Creating a Submission (As coordinator)

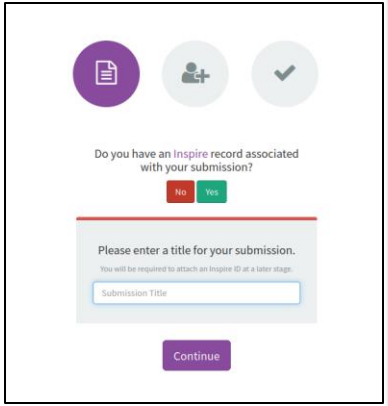
Do you have an inspire record?



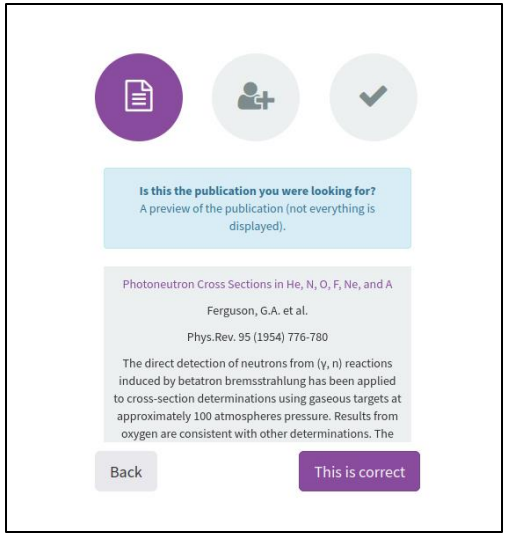
Yes: Enter Inspire ID
No: Entire title

NOTE: An Inspire ID is required to finalise the submission.



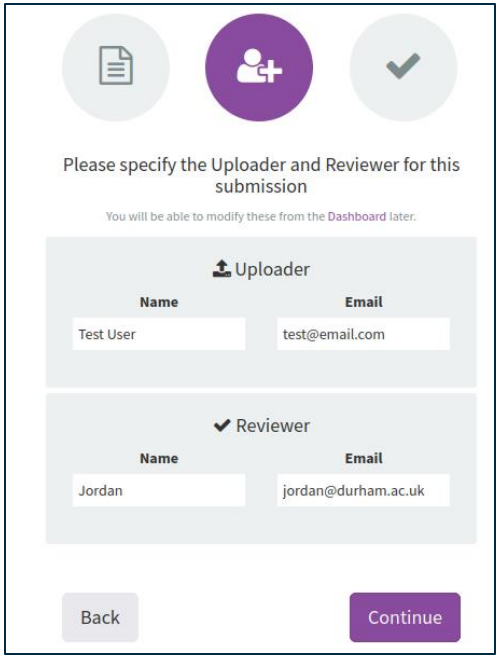


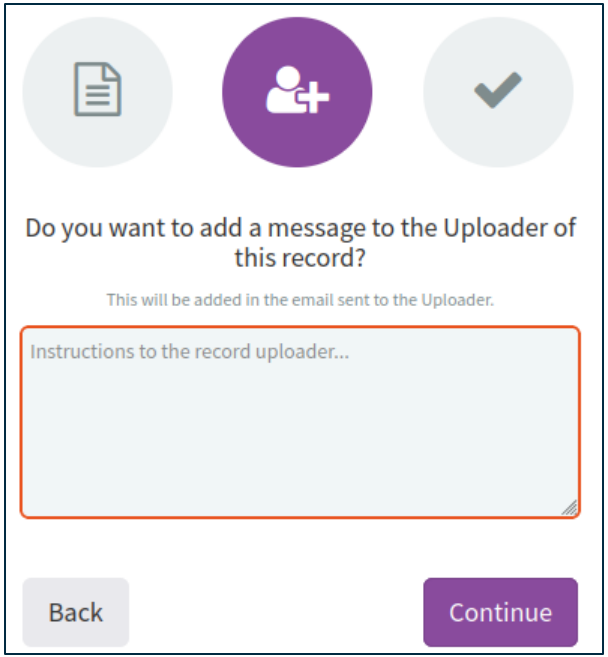
Yes: Confirm publication
No: Continues

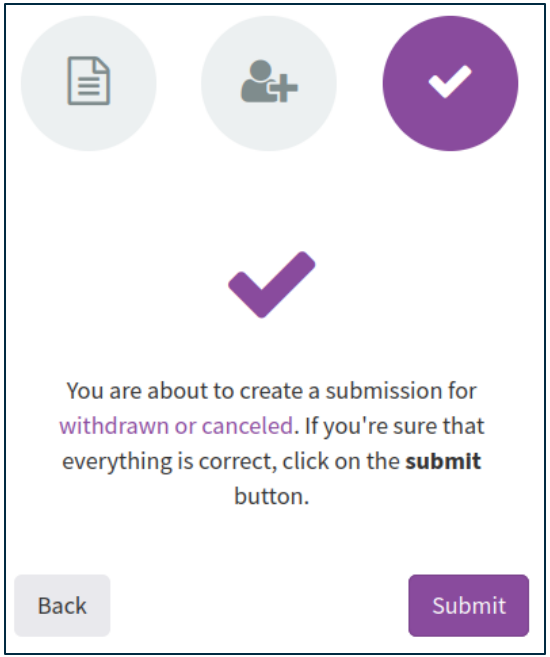


Durham University

HEPData – Creating a Submission (As coordinator)



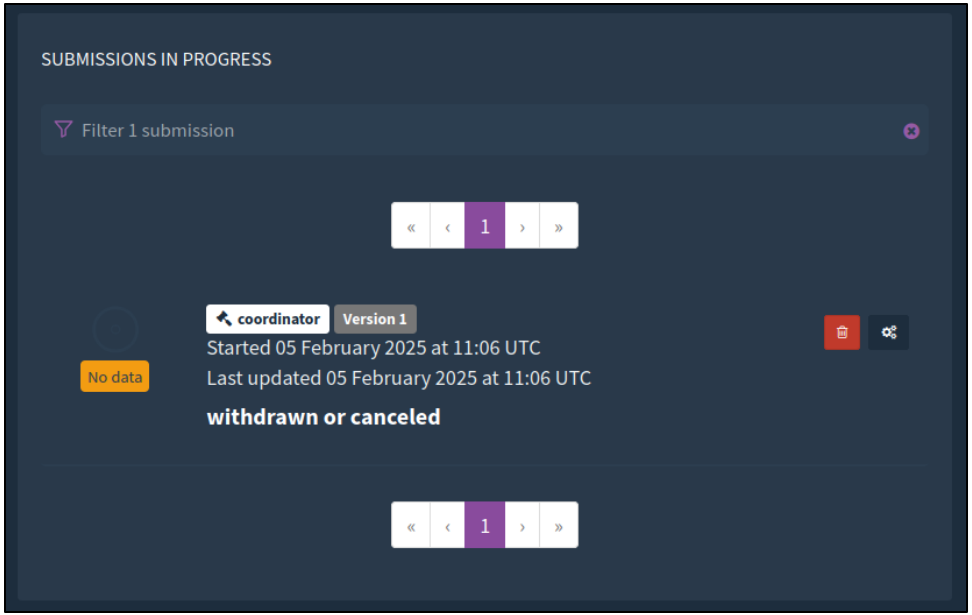




Durham University

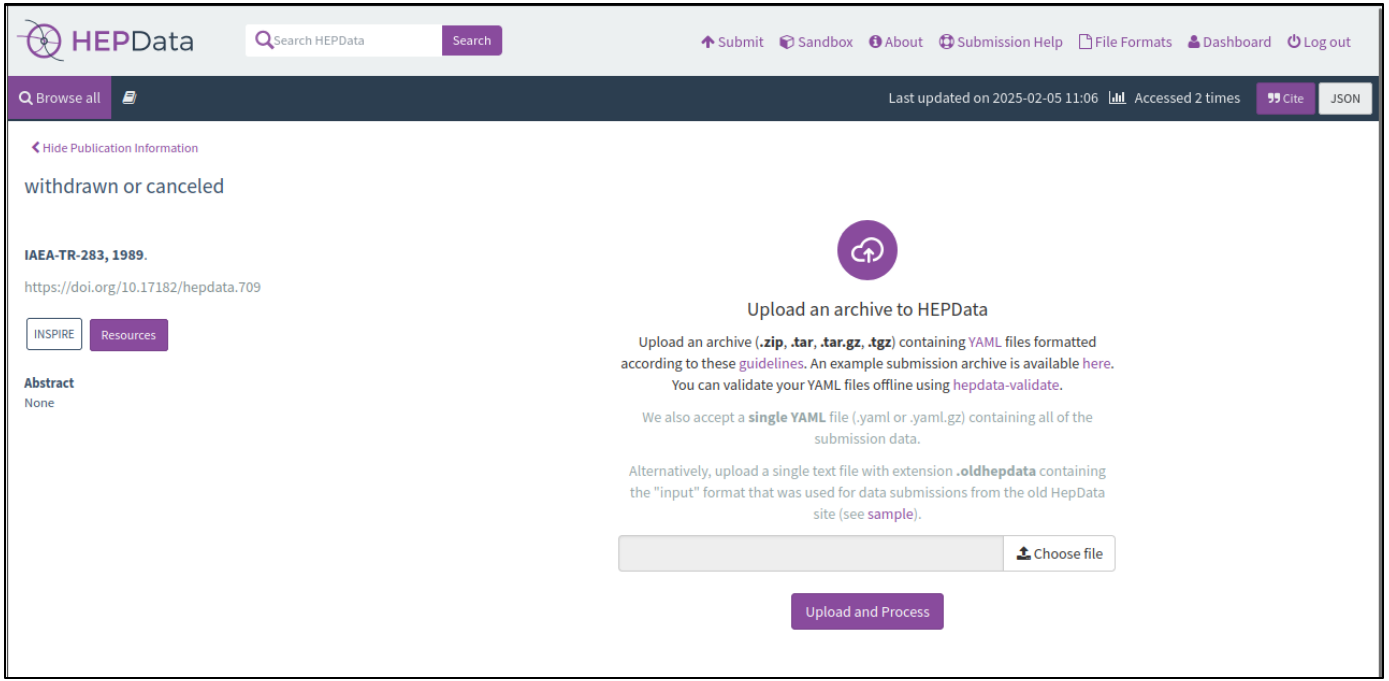
HEPData – Creating a Submission

Submission now available in the dashboard



Durham University

HEPData – Creating a Submission



Durham University

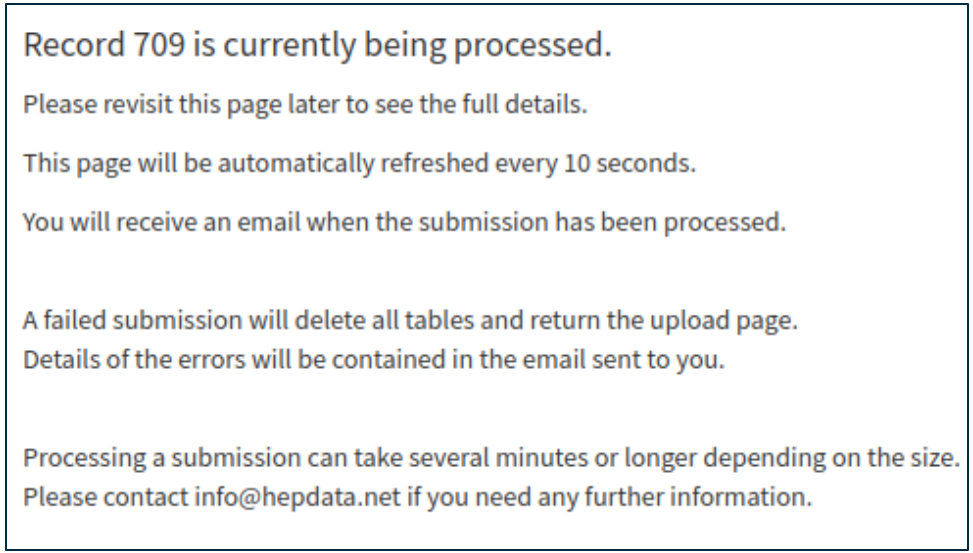
HEPData – Creating a Submission

Submission will then be processed.

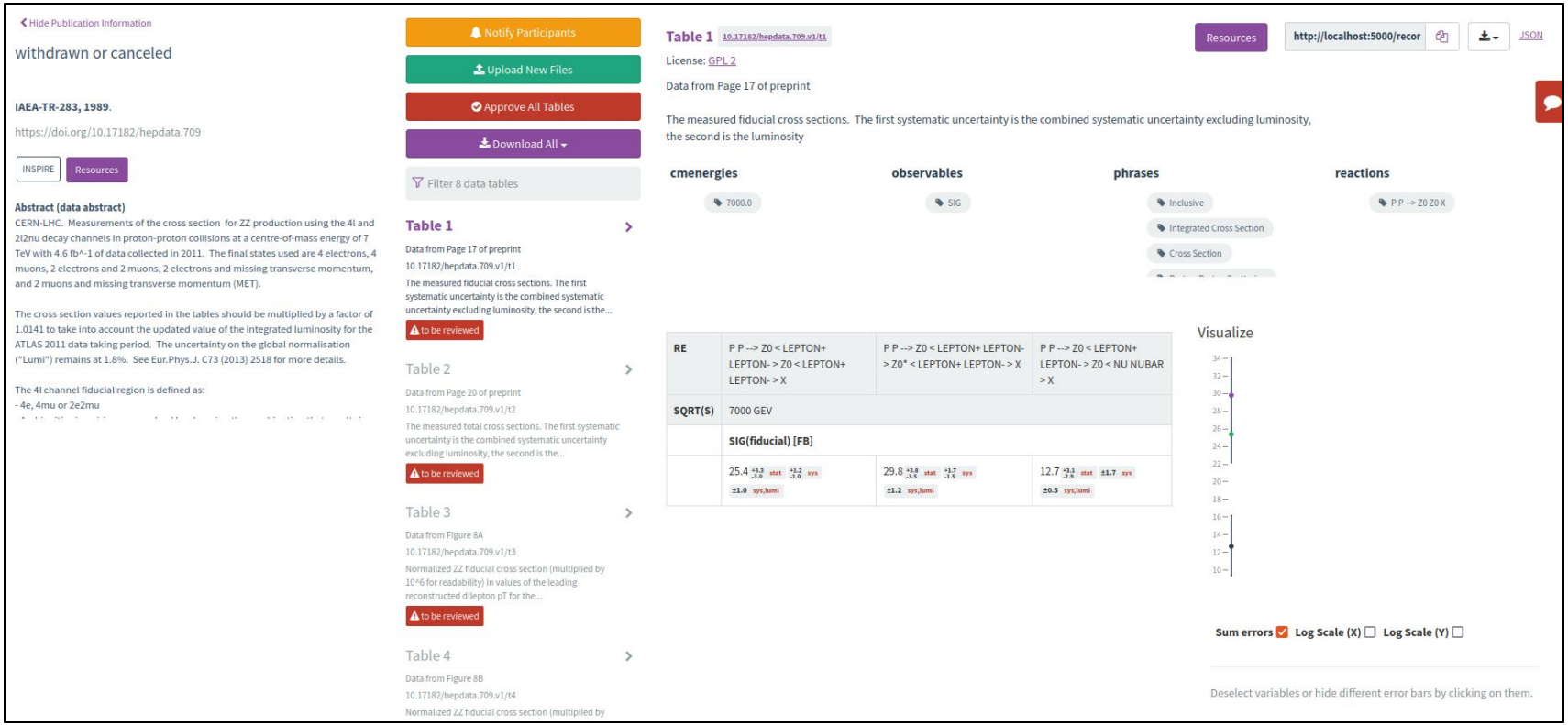
You will receive an email confirming success or failure status after data validation.

Page will reload, displaying the record, or the upload screen again.

Schema available at: https://github.com/HEPData/hepdata-validator/blob/main/hepdata_validator/schemas/1.1.1/submission_schema.json



HEPData – Creating a Submission

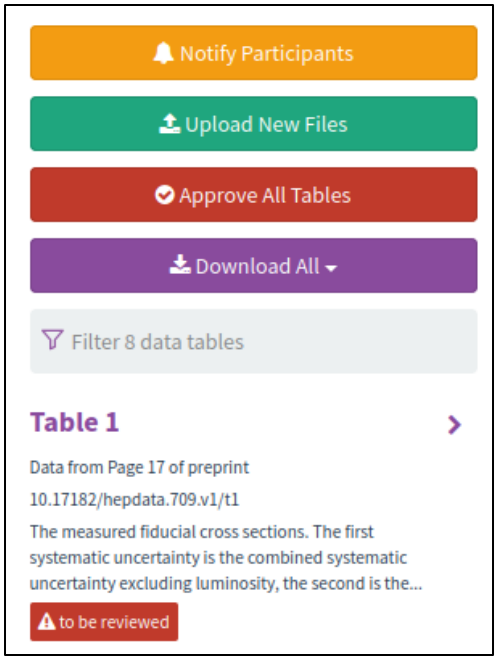


Durham University

HEPData – Submission Reviewing

Review tools:

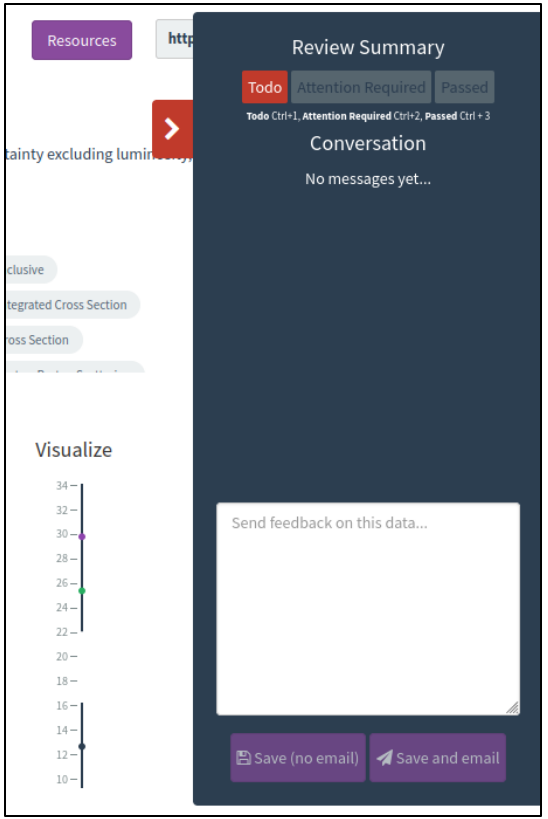
- Participant notification
- File reupload
- Mass/single table approval
- File download



Durham University

HEPData – Submission Reviewing

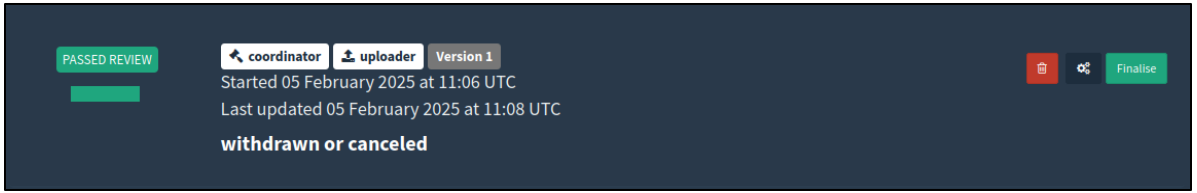
Review feedback and communication with reviewers through ***Review Summary*** widget.

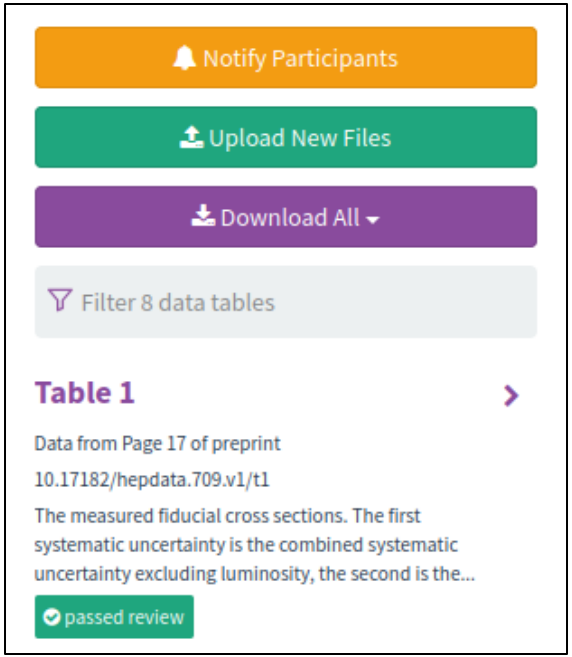


Durham University

HEPData – Submission Review

After approval, there will now be a *finalize* button!
(You must assign an Inspire ID)





Your data is now linkable, accessible and searchable.

Durham University

HEPData – GitHub

Located at: https://github.com/HEPData

- **hepdata:** Main web application
- **hepdata-validator:** JSON schema and validation code
- **hepdata-submission:** Documentation and examples
- **hepdata-converter:** YAML to CSV/ROOV/YODA converter
- **hepdata_lib:** Helps transform test/ROOT files to YAML
- **hepdata-cli:** Search/download/upload from CLI or API
- **miscellaneous:** Jupyter notebooks for various insights

Thank you!

Contact - info@hepdata.net

https://www.hepdata.net/

https://hepdata-submission.readthedocs.io/